DOCUMENT RESUME

ED 143 676 TM 006 380

AUTHOR Kraemer, Helena Chmura
TITLE A Central t-Approximation to the Non-Central t-Distribution. Technical Report Series No. 107.
INSTITUTION Stanford Univ., Calif. Dept. of Psychiatry and Behavioral Science.
NOTE 14p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Correlation; Hypothesis Testing; *Student Distribution
IDENTIFIERS *Noncentral t Distribution

ABSTRACT

There are a variety of statistical problems which involve the use of the non-central t-distribution. The development of tables for the non-central t-distribution has been hindered by the number of parameters involved. Several approaches to this task have been made in the literature, characterized by either focusing attention on the subset of possible parameter values. This paper addresses the problem of the non-central t-distribution based on the central t-distribution. The mathematical development of the approach is presented, along with advantages and uses of the approach and a table of values. (Author/JKS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED143676

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A CENTRAL t-APPROXIMATION TO THE NON-CENTRAL t-DISTRIBUTION

Helena Chmura Kraemer

Laboratory of Stress and Conflict

Technical Report Series #107

TM006 380

TECHNICAL REPORT SERIES
LABORATORY OF STRESS AND CONFLICT
Department of Psychiatry and Behavioral Science
Stanford University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Helena Chmura Kraemer
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND THE ERIC SYSTEM CONTRACTORS."

Summary

A central t-distribution approximation to the non-central t-distribution is suggested which is quite accurate even for moderate values of degrees of freedom, and is a more accurate approximation than is the normal approximation for small values of the non-centrality parameter δ.

1. Introduction and Summary

Let the parameters f, x, δ and ϵ be defined by the relationship:

$$\Pr\{t'_f(\delta) \geq x\} = \epsilon \,, \tag{1}$$

where $t'_f(\delta)$ is a random variable having a non-central t-distribution with non-centrality parameter δ and degrees of freedom, f. To deal with the broad range of statistical problems involving use of the non-central t-distribution, (see e.g., problems described in [2], [8]) one must be able to calculate any one of the parameters (f, δ, x, ϵ) as a function of the other three. In the absence of closed mathematical expressions, required would be four triple-entry tables with x and δ taking on any real value (although considerations of symmetry allow restriction of δ to non-negative real values without loss of generality), f taking on positive integer values, ϵ ranging over the unit interval.

Several approaches to this task have been made, characterized by (1) focusing attention on the calculation of only one or two of the four parameters, and (2) restriction to a subset of possible parameter values. In each case, the form of the tables and the restriction of parameter values reflect the author's focus on specific practical problems.

Thus Neyman and Tokarska ([5]) chose only two values of x for each value of f, integer values of f from 1 to 20, ϵ values: .01, .05, .1(.1).9. Within these parameter limitations, δ is tabled as a function of f, x, ϵ. Johnson and Welch ([2]) chose 17 values of ϵ but only $f = 4, 5, 6, 7, 8, 9, 16, 36, 144$ and presented tables directed

to ready calculationof δ or x as a function of the other parameters. Patnaik ([6]) chose $f = 4(1)15,20$, $\delta = 0(.5)\ 4.0$, $\epsilon = .05$, $.01$ and tabled x as a function of these. Resnikoff and Lieberman ([8]) tabled the entire cumulative probability distributions (x,ϵ) for $f = 1(1),24$, 29, 34, 39, 44, 49, but only 7 values of δ for each f, the smallest value being $.674490\ \sqrt{f+1}$.

In addition to these exact tables, there is a normal approximation to the non-central t-distribution ([2]). The relationship among the parameters is expressed in the equations:

$$\Pr\{t'_p(\delta) \leq x\} \approx \Phi((x-\delta)\ (1 + x^2/2f)^{-1/2}) \qquad (2)$$

where, if $z \sim N(0,1)$:

$$\Phi(K_\epsilon) = \Pr\{z \leq K_\epsilon\} = \epsilon \quad . \qquad (3)$$

As noted by Johnson and Welch ([2]), this approximation, though based on asymptotic theory, is quite good, even for f of moderate size.

The purpose of this paper is to add to the tools available to deal with the non-central t-distribution, another approximation, one based on the central t-distribution.

This approximation complements existing techniques in the sense that it is most accurate for parameter values for which tables of exact distributions are most sparse, and is more accurate than the normal approximation for those parameter values. It also provides an interesting

theoretical link between the non-central t-distribution and the non-null distribution of a correlation coefficient.

The relationship among the parameters based on this procedure is expressed in the equations:

$$\Pr\{t'_f(\delta) \leq x\} = \psi_f(f^{1/2}(X-\rho)\ [(1-X^2)(1-\rho^2)]^{-1/2}) \qquad (4)$$

where

$$X = x(x^2+f)^{-1/2} ,$$

and

$$\rho = \delta(\delta^2+f)^{-1/2}$$

or equivalently:

$$\Pr\{t'_f(\delta) \leq x\} = \psi_f(x(1+\delta^2/f)^{1/2} - \delta(1+x^2/f)^{1/2}) . \qquad (5)$$

Here

$$\psi_f(K_{\epsilon,f}) = \Pr\{t'_f(0) \leq K_{\epsilon,f}\} = \epsilon . \qquad (6)$$

2. Theory

Theorem: If $t'_f(\delta)$ has a non-central distribution with f degrees of freedom and non-centrality parameter δ, then there exists a function $\rho = \rho(\delta, f)$ such that

$$\lim_{f \to \infty} \left[\Pr\{t'_f(\delta) \le x\} - \Psi_f\left(f^{1/2}(X-\rho) \left[(1-X^2)(1-\rho^2)\right]^{-1/2} \right) \right] = 0, \quad (7)$$

where

$$X = x(x^2+f)^{-1/2} .$$

Proof: Clearly when $\delta = 0$, we let $\rho(\delta, f) = 0$, and the result is exact.

The exact density function of $t = t'_f(\delta)$ is (e.g., [8]):

$$\frac{\exp[-f\delta^2/2(f+t^2)]\ [f/(f+t^2)]^{(f+1)/2}}{2^{(f-1)/2}\ \Gamma(f/2)\ \Gamma(1/2)\ f^{1/2}} \int_0^\infty x^f \exp[-(x+y)^2/2]\ dx, \quad (8)$$

where $y = -\delta t (f+t^2)^{-1/2}$.

Let $r = t(t^2+f)^{-1/2}$. Thus the exact density function of r is:

$$g_f(r,\delta) = \frac{\exp[-\delta^2(1-r^2)/2]\ (1-r^2)^{(f/2)-1}}{2^{(f-1)/2}\ \Gamma(f/2)\ \Gamma(1/2)} \int_0^\infty x^f \exp[-(x-\delta r)^2/2]dx. \quad (9)$$

Since $g_f(r,\delta) = g_f(-r, -\delta)$, we can restrict attention to $\delta \ge 0$. Now $g_f(r,\delta)$ may be written

$$g_f(r,\delta) = h_f(r,\delta) \times [e^{-\delta^2/2(1-\rho^2)-f/2}]$$

$$\times [(1-\rho r)^f \exp(f^{1/2}\ \delta r)] \times E[\exp(\delta r(u^{1/2} - f^{1/2}))]. \quad (10)$$

Here

$$u \sim \chi^2_{f+1},$$

$$h_f(r,\delta) = \frac{(1-\rho^2)^{f/2}\ (1-r^2)^{(f/2)-1}}{B(f/2,\ 1/2)\ (1-\rho r)^f}, \quad (11)$$

the probability density function of the intraclass correlation coefficient r based on a sample of $(f+1)$ observations from a bivariate normal distribution with equal variances and correlation coefficient ρ. If r has density function (11) then ([3]):

$$f^{1/2}(r-\rho)\ ((1-r^2)(1-\rho^2))^{-1/2} \sim t_f'(0) \ . \tag{12}$$

Let $\rho = \rho(\delta, f)$ be a function such that $\lim_{f \to \infty} f^{1/2}\ \rho(\delta, f) = \delta$. Here for convenience let us choose specifically

$$\rho = \delta f^{-1/2} \ .$$

On this basis, each of the three bracketed factors of (10) converge uniformly for $|r| \leq 1$ to [illegible]. Thus if

$$G_f(X,\delta) = \int_{-1}^{X} g_f(r,\delta)dr = \Pr\{t_f'(\delta) \leq x\} \ , \tag{13}$$

$$H_f(X,\delta) = \int_{-1}^{X} h_f(r,\delta)dr = \psi_f(f^{1/2}(X-\rho)\ [(1-X^2)(1-\rho^2)]^{-1/2}) \ , \tag{14}$$

then

$\lim_{f \to \infty} (G_f(X,\delta) - H_f(X,\delta)) = 0$, and the theorem is proved.

An alternative way of stating the result of the theorem is that if t has a non-central t-distribution with degrees of freedom f and non-centrality prarmeter δ, then

$$r = t(t^2+f)^{-1/2}$$

has approximately the distribution of the intraclass correlation coefficient with f degrees of freedom and $\rho = \delta(\delta^2+f)^{-1/2}$, i.e., (11). Thus procedures designed to define tests of hypothesis about ρ, to calculate the power of such tests, to compute confidence intervals for ρ, and to test homogeneity of ρ_i, $i = 1, 2, \ldots, K$, in K different populations ([4]) can be directly used to perform, at least approximately, similar tasks concerning the non-centrality parameter δ. Thus, for example, a 95% one sided confidence interval for ρ based on r is

$$\rho \geq \frac{r-c}{1-rc} = L \,, \qquad \text{where } r = t(t^2+f)^{-1/2} \,, \tag{15}$$

and c is the 95-th percentile of the distribution of intraclass correlation coefficient when $\rho = 0$, degrees of freedom = f (tabled, for example, in Table 13, [7]). The 95% one-sided confidence interval for δ is:

$$\delta \geq \frac{\sqrt{f}\, L}{\sqrt{1-L^2}} = \frac{\sqrt{f}\,(r-c)}{\sqrt{(1-r^2)(1-c^2)}} \,. \tag{16}$$

3. Comparison of the Normal- and t-Approximations

The only general case in which one or the other of these approximations is unequivocally preferred is when $\delta = 0$ and trivially the t-approximation, being exact, yields more accurate results. Furthermore, when f is large relative to x and δ, then

$$\Phi((x-\delta)\ (1+x^2/2f)^{-1/2}) \approx \Phi_f(x(1+\delta^2/f)^{1/2} - \delta(1+x^2/f)^{1/2}) \ .$$

In this case the major practical difference between the approximations will be that of convenience rather than accuracy. For f of moderate size and non-zero non-centrality paramter δ, neither approximation is uniformly better.

The accuracy of the t-approximation depends on the particular choice of the function $c(\delta,f)$. We have found empirically that $c(\delta,f) = \delta(\delta^2+f)^{-1/2}$ seems to yield the best results in terms of approximating the tail probabilities of the non-central t-distribution. For this choice, both approximations yield δ as the estimate of the median of the $t'_f(\delta)$, whereas the true median exceeds δ. Thus for x near δ or probability near .5, the two approximation procedures perform equally.

To compare the approximations more generally, we note that for $\delta > 0$,

$$\begin{aligned} \Phi(z) < \Pr\{t'_f(\delta) \le x\} \qquad & x < 0 \\ = \Pr\{t'_f(\delta) \le x\} \qquad & x = 0 \\ > \Pr\{t'_f(\delta) \le x\} \qquad & x > 0 \ , \end{aligned} \tag{17}$$

where $z = (x-\delta)\ (1 + x^2/2f)^{-1/2}$ ([1]), and

$$\Phi_f(t) > \Pr(t'_f(\delta) \leq x) \qquad x < e_{f,\delta} \qquad (13)$$

$$= \Pr(t'_f(\delta) \leq x) \qquad x = e_{f,\delta}$$

$$< \Pr(t'_f(\delta) \leq x) \qquad x > e_{f,\delta}$$

where $t = x(1 + \delta^2/f)^{1/2} - \delta(1 + x^2/f)^{1/2}$ and $e_{f,\delta}$ is a point depending on f and δ, with $\delta < e_{f,\delta}$. Thus one-tailed tests and confidence intervals based on the t-approximation will tend to be on the conservative side; the normal approximation will tend to err on the liberal side.

The errors are not, however, balanced, since it can be shown by empirical study that for large δ, the t-approximation error though conservative will be substantial while the normal-approximation error though liberal will be quite small. This point is illustrated in Table 1 in which the parameter λ tabled by Johnson and Welch ([2]), used to compute exact values of the 95-th percentile of the $t'_f(\delta)$ distribution, is compared to the values of λ as derived from the t-approximations for $f = 4, 36$. The normal approximation would yield $\lambda = 1.6449$ for all f, δ. For the task of estimating the 95-th percentile, one would prefer the t-approximation for $\delta < 2$, and would tend to prefer the normal-

[Table 1 about here]

approximation for larger δ. The criterion of what constitutes large or small δ would vary depending on the task, but the general rule would not: the t-approximation is preferred for small δ, the normal-approximation for large δ.

References

[1] Hawkins, D.M. (1975). "From the Noncentral t to the Normal Integral." The American Statistician, 29, 42-43.

[2] Johnson, N.L., and Welch, B.L. (1940), "Applications of the Non-Central t-Distribution." Biometrika, 31, 362-389.

[3] Kraemer, H.C. (1973). "Improved Approximation to the Non-Null Distribution of the Correlation Coefficient." J. Amer. Statist. Assoc., 68, 1004-1008.

[4] Kraemer, Helena Chmura (1976). "On Estimation and Hypothesis Testing Problems for Correlation Coefficients." Psychometrika, 40, 473-485.

[5] Neyman, J., and Tokarska, B. (1936), "Errors of the Second Kind in Testing "Student's" Hypothesis." J. Amer. Statist. Assoc., 31, 318-326.

[6] Patnaik, P.B. (1955). "Hypothesis Concerning the Means of Observations in Normal Samples." Sankhya, 15, 343-372.

[7] Pearson, E.S. and Hartley, H.O. (1962). Biometrika Tables for Statisticians (Vol. I). London: Cambridge University Press.

[8] Resnikoff, G.J., and Lieberman, G.J. (1957). Tables of the Non-Central t-Distribution. Stanford University Press, Stanford, California.

Table 1:

Exact Values of $\lambda^{3.}$ as a function of δ, f where the 95-th percentile of the $t'_f(\delta)$ distribution is

$$[\delta+\lambda(1+\delta^2/2f-\lambda^2/2f)^{1/2}]\,[1-\lambda^2/2f]^{-1}$$

compared to those based on the central t-approximation λ_t.

Based on the normal approximation $\lambda_n = 1.6449$ for all f, δ.

f=degrees of freedom		4			36		
η'[1]	ρ[2]	δ	λ	λ_t	δ	λ	λ_t
.9	.9461	5.84	1.643	1.707	17.52	1.667	2.008
.8	.8835	3.77	1.650	1.703	11.32	1.669	1.956
.7	.8110	2.77	1.657	1.699	8.32	1.671	1.905
.6	.7276	2.12	1.664	1.695	6.36	1.671	1.850
.5	.6325	1.63	1.671	1.691	4.90	1.671	1.800
.4	.5252	1.23	1.679	1.688	3.70	1.670	1.754
.3	.4064	.889	1.687	1.686	2.67	1.668	1.714
.2	.2774	.557	1.693	1.687	1.73	1.665	1.683
.1	.1407	.284	1.698	1.692	.853	1.661	1.663
0.0	.0000	.000	1.703	1.703	.000	1.656	1.656

1. $\eta' = \delta(2f + \delta^2)^{-1/2}$

2. $\rho = \delta(\delta^2 + f)^{-1/2} = 2^{1/2}\,\eta'(1 + \eta'^2)^{-1/2}$

3. From Table V in [2].

List of Symbols

δ = delta

ϵ = epsilon

ρ = rho

Γ = capital gamma

Ψ = capital psi

Φ = capital phi

λ = lambda